

Ensemble learning model for Wifi indoor positioning systems

Pham Doan Tinh, Ta Thi Ngoc Mai

Department of Electronics and Computer Engineering, Hanoi University of Science and Technology, Vietnam

Article Info

Article history:

Received Aug 24, 2020

Revised Jan 19, 2021

Accepted Feb 20, 2021

Keywords:

Ensemble learning

Indoor

Positioning

WiFi RSSI fingerprinting

ABSTRACT

WiFi indoor positioning researches have received much attention from researchers recently. In this research, we focus on studying the performance of indoor positioning systems that utilize our new proposed ensemble machine learning model. Our new ensemble learning model uses several models for normal data training and position prediction, then it uses the verification data together with its' prediction errors from trained models as the input data to train an intermediate classification model to classify which set of Wifi received signal strength indicator (RSSI) is the best match for each position prediction model. The experimental result shows that our proposed ensemble model outperforms other compared models.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Pham Doan Tinh

Department of Electronics and Computer Engineering

School of Electronics and Telecommunications

Hanoi University of Science and Technology

No. 1 Dai Co Viet street, Hanoi, Vietnam

Email: tinh.phamdoan@hust.edu.vn

1. INTRODUCTION

Location-aware services need position information to carry out a specific task. In outdoor environments, one can use the global positioning system (GPS) equipped devices to get position information, but it is hard to use GPS in indoor environments. Recent researches have shown that WiFi indoor positioning systems (WiFi-IPSs) are very promising for those services. Many Wifi-IPS algorithms and systems have been proposed so far, but we can categorize them as range-based and range-free algorithms [1]. The WiFi RSSI fingerprinting systems (or range-free systems) that use WiFi signals from the surrounding wireless access points (APs) to provide the object's location information. This method eases the deployment at a low cost, and they require no extra infrastructure. Researches have shown that WiFi fingerprinting technology using the received signal strengths received from WiFi access points is a very promising method for indoor positioning [2]. However, this method causes many difficulties as Wi-Fi RSSI suffers from multipath and shadowing interferences in indoor dynamic environments [3].

Therefore, the measured RSSI value is not stable and highly depends on the measuring environment and surrounding objects. RSSI positioning estimates also have relatively low accuracy and security [4]. In particular, when predicting the movement trajectory of a person or device indoor environment, the more mobility, the bigger error is. Many methods have been proposed to overcome this limitation. For example, the use of the average of many selected maximum RSSI observations [5]. It uses a smoothness index to test the quality of RSSI to select an appropriate number of RSSI observations. A multi-point fingerprint matching algorithm has also been proposed, in which the common single-point matching procedure is expanded into multiple points [6].

Traditional ensemble methods combine the predictions from several base estimators to improve the generalizability and robustness over a single estimator. In averaging methods, the principle is to build several estimators independently and then to average their predictions. In contrast, the boosting methods, base estimators are built sequentially and one tries to reduce the bias of the combined estimator. Singh *et al.* [7] propose a method using an ensemble of classifiers on weighted averages of WiFi RSSI values within a time window to localize a user in an apartment. Akram *et al.* [8] combine gaussian mixture model (GMM)-based soft clustering and random decision forest (RDF) ensembles on WiFi fingerprints to solve the indoor localization at both room-level and latitude-longitude prediction. Lee *et al.* [9] employ random forest ensemble learning method on WiFi RSSI to locate the location of a user. Some authors try to employ the particle swarm optimization together with ensemble learning to solve the indoor localization using ultra-wide-band signals (UWB) [10]. Some researchers try to make modification to traditional ensemble learning models or combine them in different ways to find out position of a user in indoor environments based on WiFi RSSI signals [11-16].

This research proposes a new ensemble model. In this model, an intermediate classifier is used to select the best model for each test point and its prediction from base models. The best model is defined as the model that gives the smallest prediction error. This allows the new model to choose the best estimator for each test data point flexibly.

2. RESEARCH METHOD

2.1. WiFi fingerprint positioning

Suppose that each location has a particular set of WiFi signal strength from APs measured in the offline phase, called an offline fingerprint, then the fingerprint got during the online phase is compared to the offline fingerprints stored in the database to estimate the position of an object. In the offline phase, each reference point includes signal strength measured from all accessible APs together with known 2D coordinates. When an object enters the region, they compare its current measured RSSI data and stored offline data to infer the position of the object [2, 17-20].

Wi-Fi RSSI Fingerprint data from surrounding access points formed a map for an area with some probability distribution of RSSI values at each given location (x, y). In most methods, RSSI values of the online phase are compared with those stored in the database in the surveying phase to find the closest match, then the position (x, y) is predicted based on this match [21]. A fingerprint is a set of signal strengths from surrounding access points over time at a given location (called reference point). These fingerprints have some relation to locations associated with them, so they can be used to distinguish those locations. When applying machine learning models to this kind of problem, most methods perform two separate phases. In the first phase, the training phase, multiple WIFI RSSI fingerprints scanned at each reference point together with its coordinates are used to train the learning model, and they are also recorded to the database for future uses. In the second online phase, it forwards a new scanned RSSI fingerprint at the unknown location to the trained model to predict the unknown position (x, y). The most commonly used estimation method is the K-Nearest Neighbor (KNN). More complicated methods include support vector machine (SVM), deep neural networks (DNN), the hidden markov model (HMM), and Gaussian Process Assisted are also implemented [2, 22-25].

Now, let us formalize the methods using a mathematical model. Assume that, the signal strengths at N points are measured together with their coordinates, respectively in a room (maybe a squared division or randomized). Normally, people will use 80% of that data set for training and the remaining 20% is used for testing purposes. Each point P_i in Figure 1 has the corresponding coordinate (x_i, y_i) and a set of RSSI values from access points: $P_i(x_i, y_i, \text{RSSI}_i)$, with $\text{RSSI}_i = (r_1^i, r_2^i, \dots, r_M^i)$ are RSSI values from M access points and r_j^i ($j=1..M$) is the RSSI of the j^{th} access point at the point i. The Euclidean distance from point P_i to any point P_j is calculated by the formula:

$$d_{ij} = \sqrt{(r_1^i - r_1^j)^2 + \dots + (r_M^i - r_M^j)^2} \quad (1)$$

The point P_i' coordinates are predicted based on the K nearest points. More complicated algorithms such as deep neural network (DNN), support vector regression (SVR), and other modern learning models can be applied to give better accuracy but, we have to pay more intensive computation.

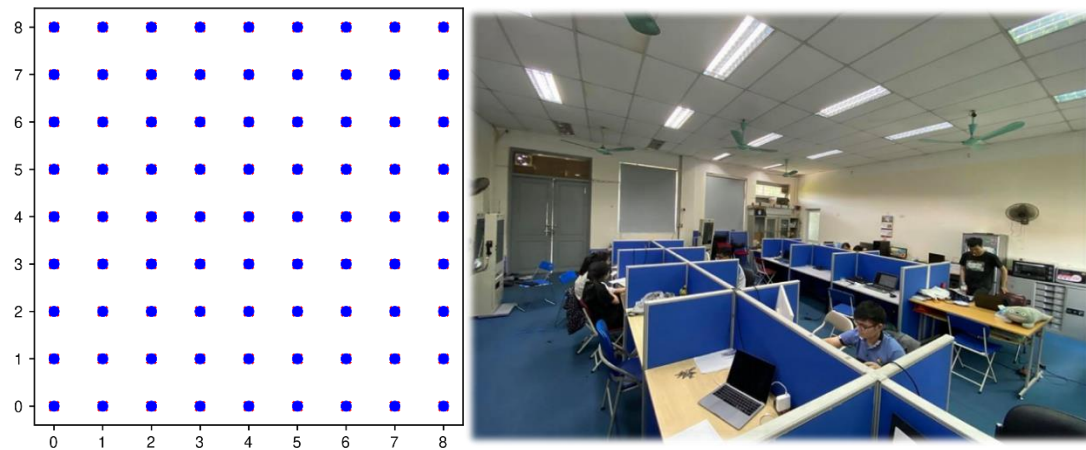


Figure 1. Actual experimental environment and the setup for training data points

2.2. Proposed ensemble method

After training base models (KNN, DNN, and random forest (RF)), we use these models to predict test points' coordinates, we realize that prediction errors of a single point from different models are also different. Which means that a model is good for a subset of test points. Therefore, we may think about building an intermediate classifier to classify which a point is the best fit with a specific base model (KNN, DNN, and RF). Our proposed ensemble model is illustrated in Figure 2.

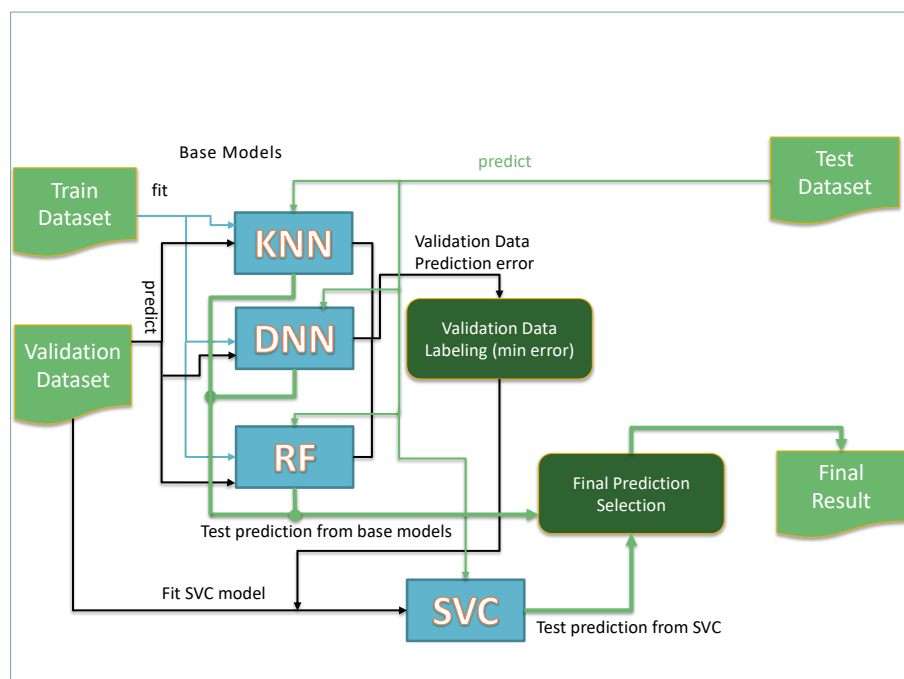


Figure 2. The proposed new ensemble model

Specifically, we divide the RSSI dataset measured in the offline phase into three different data sets: training, validation, and test sets. The training set is used to train three base models. The three trained base models are used to predict the validation set. The validation prediction errors and the corresponding best fit models are then used to train the support vector classifier (SVC) as an intermediate classifier (each data point is labeled with an integer that represents a base model index in the base model list). Each point in the test set

is forwarded to the trained SVC model to find an appropriate base model for the final prediction for the test point. In short, our new ensemble works with the following steps:

1. Step 1: New ensemble model initialization
 - Initialize base regression machine learning models
 - Initialize the SVC intermediate classification model
2. Step 2: Training new ensemble model
 - Use the training data set to train for base regression models
 - Make prediction on the validation data set from trained base models and evaluate the errors
 - Label each data point in the validation set with corresponding base model that gives smallest error for that validation point
 - Validation data set together with base model labels are used to train the intermediate SVC model
3. Step 3: New ensemble prediction
 - Forward each test data input into the SVC model to get the best model from the intermediate classifier
 - Use the selected best base model to make prediction and get the final best predicted coordinates for each test data point

2.3. Validation data labeling

Table 1 shows some sample validation data points used to forward to base models. The corresponding minimum error is used to label the validation data that are used to train the intermediate classifier (SVC).

Table 1. Data labeling for intermediate classifier

Validation Data Points	Model Prediction Error (m)			Label (Index of the model giving minimum error)
	KNN (0)	DNN (1)	RF (2)	
1	1.99	2.97	2.43	0
2	1.00	0.47	1.78	1
3	0.53	1.08	1.12	0
4	2.36	4.90	1.35	2
5	2.89	1.70	2.13	1
6	1.43	0.46	0.50	1
7	0.84	0.75	0.74	2
8	1.19	1.58	0.82	2
9	2.86	4.30	3.16	0
10	2.10	1.15	1.85	1
...				

3. RESULTS AND DISCUSSION

The experiments are conducted in our lab room with a dimension of 10 by 10 meters, and the room is divided into grids of 9 by 9 points to measure the WiFi RSSI fingerprints of surrounding access points in the surveying phase. The room contains tables, chairs, computers, and other networking devices as well as human beings working in the room. In this specific scenario, we use RSSI values from 9 access points surrounding the room and nearby rooms (there are about 600 measured points). Validation dataset is extracted from the measured training points and they are excluded from training dataset. For the testing data, we further measure RSSI fingerprints of random points in the room. To evaluate the performance of our proposed algorithms, the Euclidean distance between the estimated and true location is used to measure the error. Let (x_i, y_i) be the true 2-D physical coordinates and (\hat{x}_i, \hat{y}_i) be the estimated location of point P_i , respectively, then the distance error Err_d is computed:

$$Err_d = \frac{\sum_{i=1}^N \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2}}{N} \quad (2)$$

For the experiments, we use python sklearn libraries with three base models with the hyper-parameter configurations are: RandomForestRegressor (random_state=0, n_estimators=300), KNeighborsRegressor (n_neighbors=7, weights='distance'), and MLPRegressor (hidden_layer_sizes = (128,64,32,16), activation='relu', solver='adam', batch_size=3, max_iter=300, random_state=0). For the intermediate classifier, we use SVC (random_state=0). The VotingRegressor model uses the same three base models as in our proposed ensemble model.

Figure 3 shows the mean error for each base model and the new ensemble model (the data is also showed in Table 2). From the figure, one can easily find that the new ensemble model gets very good accuracy in comparison to each base model used. Figure 4 represents numerical analysis about the base

model selection (the data is also showed in Table 3). For each base model, the percentage number shows the rate at which a model gives the smallest error among three base models on the test data. The percentage number of the new ensemble model shows the rate that the intermediate classification model (SVC) correctly classifies the base model which gives the smallest error. In this specific scenario, it achieves 60.38% ratio (higher than that of each base model's).

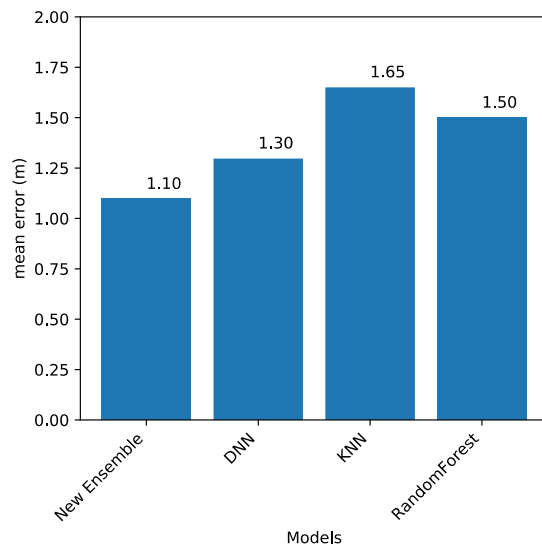


Figure 3. Mean error of base models and our proposed ensemble model

Table 2. Mean error of base models and our proposed ensemble model

Model	Mean Error (m)
New Ensemble	1.10
DNN	1.30
KNN	1.65
RandomForest	1.50

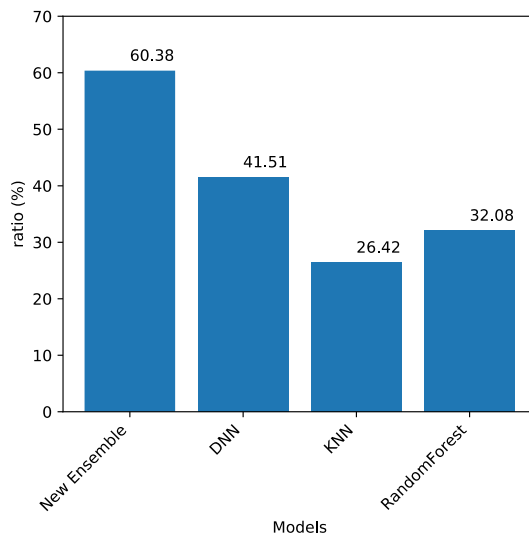


Figure 4. The ratio rate (%) at which a model correctly gives the minimum error

Table 3. The ratio rate (%) at which a model correctly gives the minimum error

Model	Ratio (%)
New Ensemble	60.38
DNN	41.51
KNN	26.42
RandomForest	32.08

The cumulative error distribution for each base model and the new ensemble model is illustrated in Figure 5. We also evaluated our proposed ensemble model with the other ensemble learning models such as VotingRegressor and ExtraTreeRegressor, and the comparison result is presented in Figure 6 (the data is also showed in Table 4Table 2). From the figure, it is very clear that our proposed ensemble model has very good accuracy in comparison to other ensemble learning models for this specific WiFi RSSI dataset. From the experiments, we also realized that the validation dataset should be large enough to thoroughly train the

intermediate classifier model so that the proposed ensemble model can give good prediction on the testing data.

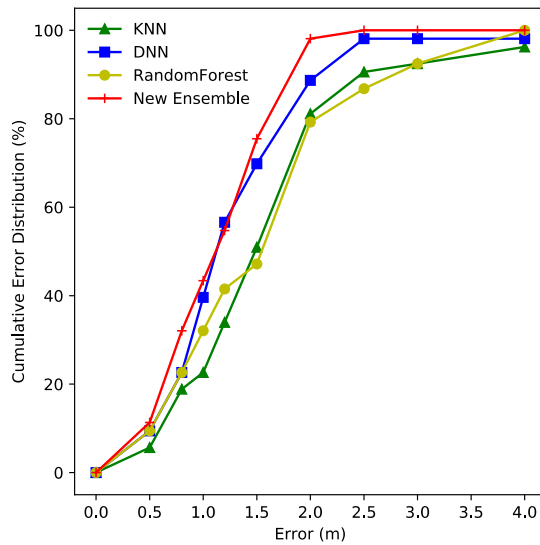


Figure 5. Error distribution of models

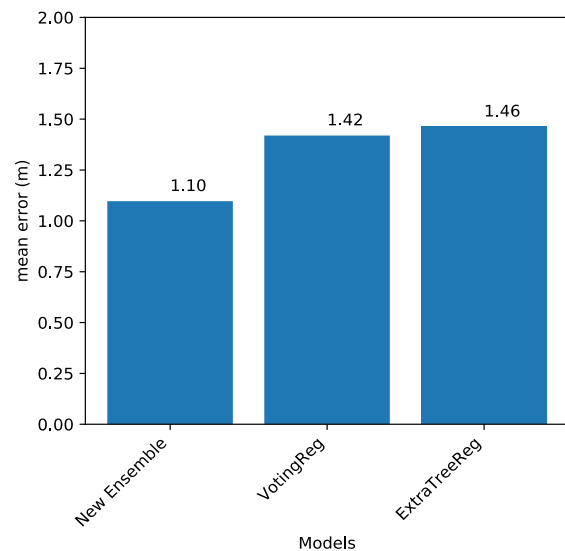


Figure 6. Mean error of ensemble models

Table 4. Mean error of ensemble models

Ensemble Model	Mean Error (m)
New Ensemble	1.10
VotingRegressor	1.42
ExtraTreeRegressor	1.46

4. CONCLUSION

The paper already analyzed our proposed new ensemble learning model. Compared to traditional ensemble models, our proposed model uses an intermediate classification model to train a validation data set, then the trained classifier is used to select the best model for each test data point. The intensive experiments have confirmed that our model has better accuracy in comparison to traditional ensemble models as well as base models.

ACKNOWLEDGEMENTS

This research is supported by the grant KC.01.13/16-20 belonging to the national program KC.01/16-20 lead by Vietnamese Ministry of Science and Technology.

REFERENCES

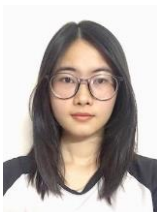
- [1] Tinh P., Kawai M., "Distributed Range-Free Localization Algorithm Based on Self-Organizing Maps," *J Wireless Com Network* 2010, 692513 (2009). <https://doi.org/10.1155/2010/692513>.
- [2] Husen, Mohd Nizam, and Sukhan Lee, "Indoor human localization with orientation using WiFi fingerprinting," *In Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication*, pp. 1-6. 2014. DOI: 10.1145/2557977.2557980.
- [3] A. Khalajmehrabadi, N. Gatsis and D. Akopian, "Modern WLAN Fingerprinting Indoor Positioning Methods and Deployment Challenges," *in IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1974-2002, thirdquarter 2017. DOI: 10.1109/COMST.2017.2671454.
- [4] Bae, Youngchul, "Robust Localization for Robot and IoT Using RSSI," *Energies* 12.11 (2019): 2212.
- [5] Xue, Weixing, *et al.* "Improved Wi-Fi RSSI measurement for indoor localization," *IEEE Sensors Journal*, vol. 17, no. 7, pp. 2224-2230, 2017. DOI: 10.1109/JSEN.2017.2660522
- [6] Li, Wen, *et al.*, "A novel method of WiFi fingerprint positioning using spatial multi-points matching," *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2016.

- [7] V. Singh, G. Aggarwal and B. V. S. Ujwal, "Ensemble based real-time indoor localization using stray WiFi signal," *2018 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, 2018, pp. 1-5, doi: 10.1109/ICCE.2018.8326317.
- [8] B. A. Akram, A. H. Akbar and O. Shafiq, "HybLoc: Hybrid Indoor Wi-Fi Localization Using Soft Clustering-Based Random Decision Forest Ensembles," in *IEEE Access*, vol. 6, pp. 38251-38272, 2018, doi: 10.1109/ACCESS.2018.2852658.
- [9] Lee, S., Moon, N. "Location recognition system using random forest", *J Ambient Intell Human Comput*, vol. 9, 1191-1196, 2018. <https://doi.org/10.1007/s12652-018-0679-5>
- [10] Xufen Cai, Long Ye, Qin Zhang, "Ensemble learning particle swarm optimization for real-time UWB indoor localization," *EURASIP Journal on Wireless Communications and Networking*, 2018.
- [11] S. Tewes, *et al.*, "Ensemble-Based Learning in Indoor Localization: A Hybrid Approach," *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, Honolulu, HI, USA, 2019, pp. 1-5, doi: 10.1109/VTCFall.2019.8891416.
- [12] Ashraf Imran, *et al.*, "DeepLocate: Smartphone Based Indoor Localization with a Deep Neural Network Ensemble Classifier," *Sensors*, vol. 20, no. 1, pp 133, 2020. <https://doi.org/10.3390/s20010133>.
- [13] X. Wang and Y. Feng, "An Ensemble Learning Algorithm for Indoor Localization," *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, Chengdu, China, 2018, pp. 774-778, doi: 10.1109/CompComm.2018.8780770.
- [14] Priya Roy, *et al.*, "Novel weighted ensemble classifier for smartphone based indoor localization", *Expert Systems with Applications*, vol. 164, 113758, 2021. <https://doi.org/10.1016/j.eswa.2020.113758>.
- [15] Krzysztof Trawinski, Jose M. Alonso, Oscar Cordon, "Applying Random Linear Oracles with Fuzzy Classifier Ensembles on WiFi Indoor Localization Problem," *Computer Science A Passion for Fuzzy Sets 2015*
- [16] Sunmin Lee, Jinah Kim, Nammee Moon, "Random forest and WiFi fingerprint-based indoor location recognition system using smart watch," *Human-centric Computing and Information Sciences*, vol 9, 2019.
- [17] S. He and S. - G. Chan, "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons," in *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466-490, Firstquarter 2016.
- [18] F. Zafari, A. Gkelias and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568-2599, thirdquarter 2019.
- [19] Simon Yiu, Marzieh Dashti, Holger Claussen, Fernando Perez-Cruz, "Wireless RSSI fingerprinting localization," *Signal Processing*, Volume 131, pp. 235-244, 2017. <https://doi.org/10.1016/j.sigpro.2016.07.005>.
- [20] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo and G. Seco-Granados, "Survey of Cellular Mobile Radio Localization Methods: From 1G to 5G," in *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 1124-1148, doi: 10.1109/COMST.2017.2785181.
- [21] Navarro, Eduardo, Benjamin Peuker, Michael Quan, Christopher Clark, and Jennifer Jipson. "Wi-Fi Localization Using RSSI Fingerprinting," PhD diss., California Polytechnic State University, 2010
- [22] Dinh-Van, N.; Nashashibi, F.; Thanh-Huong, N.; Castelli, E. Indoor Intelligent Vehicle localization using WiFi received signal strength indicator. In *Proceedings of the IEEE Mtt-S International Conference on Microwaves for Intelligent Mobility*, Nagoya, Japan, 2017, pp. 33–36.
- [23] S. Yiu and K. Yang, "Gaussian Process Assisted Fingerprinting Localization," in *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 683-690, doi: 10.1109/IIOT.2015.2481932.
- [24] LANDAU, Yael, BEN-MOSHE, Boaz. "STEPS: An Indoor Navigation Framework for Mobile Devices", *Sensors*, vol. 20, no. 14, 2020. doi: 10.3390/s20143929.
- [25] Sun, Haotai, *et al.*, "WiFi Based Fingerprinting Positioning Based on Seq2seq Model," *Sensors*, vol. 20, no. 13, 2020. Doi: 10.3390/s20133767.

BIOGRAPHIES OF AUTHORS



Tinh D. Pham is a Lecturer at the Department of Electronics and Computer Engineering, School of Electronics and Telecommunications, Hanoi University of Science and Technology, Vietnam. In 2011, he earned Doctor of Engineering Degree from Ritsumeikan University, Japan. He also obtained Bachelor and Master Degree in Electronics and Computer Engineering from Hanoi University of Science and Technology (Vietnam) in 1996 and 2000, respectively.



Mai TN. Ta is a Student in Electronics and Computer Engineering Major of Hanoi University of Science and Technology (Vietnam). Her research interests are in fields of Machine Learning Applications and Wireless Positioning.